



## OPEN ACCESS

## EDITED BY

María Sofía Urbieto,  
CONICET – UNLP, Argentina

## REVIEWED BY

Leandro Guerrero,  
CONICET Instituto de Investigaciones  
en Ingeniería Genética y Biología  
Molecular Dr. Héctor N. Torres  
(INGEBI), Argentina  
Devaki Bhaya,  
Carnegie Institution for Science (CIS),  
United States  
Roger Garrett,  
University of Copenhagen, Denmark

## \*CORRESPONDENCE

Beatriz Díez  
bdiez@bio.puc.cl

## SPECIALTY SECTION

This article was submitted to  
Extreme Microbiology,  
a section of the journal  
Frontiers in Microbiology

RECEIVED 13 October 2022

ACCEPTED 17 November 2022

PUBLISHED 02 December 2022

## CITATION

Salgado O, Guajardo-Leiva S,  
Moya-Beltrán A, Barbosa C, Ridley C,  
Tamayo-Leiva J, Quatrini R,  
Mojica FJM and Díez B (2022) Global  
phylogenomic novelty of the Cas1  
gene from hot spring microbial  
communities.  
*Front. Microbiol.* 13:1069452.  
doi: 10.3389/fmicb.2022.1069452

## COPYRIGHT

© 2022 Salgado, Guajardo-Leiva,  
Moya-Beltrán, Barbosa, Ridley,  
Tamayo-Leiva, Quatrini, Mojica and  
Díez. This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s)  
are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Global phylogenomic novelty of the Cas1 gene from hot spring microbial communities

Oscar Salgado<sup>1,2</sup>, Sergio Guajardo-Leiva<sup>3,4</sup>,  
Ana Moya-Beltrán<sup>5</sup>, Carla Barbosa<sup>1,6,7</sup>, Christina Ridley<sup>1</sup>,  
Javier Tamayo-Leiva<sup>1</sup>, Raquel Quatrini<sup>5,8</sup>,  
Francisco J. M. Mojica<sup>9</sup> and Beatriz Díez<sup>1,10,11\*</sup>

<sup>1</sup>Department of Molecular Genetics and Microbiology, Biological Sciences Faculty, Pontifical Catholic University of Chile, Santiago, Chile, <sup>2</sup>Núcleo de Ciencias Naturales y Exactas, Universidad Adventista de Chile, Chillán, Chile, <sup>3</sup>Departamento de Microbiología, Universidad de Talca, Talca, Chile, <sup>4</sup>Centro de Ecología Integrativa, Universidad de Talca, Talca, Chile, <sup>5</sup>Centro Científico y Tecnológico de Excelencia Ciencia & Vida, Santiago, Chile, <sup>6</sup>Departamento de Geología, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, Chile, <sup>7</sup>Centro de Excelencia en Geotermia de Los Andes (CEGA-Fondap), Santiago, Chile, <sup>8</sup>Facultad de Medicina y Ciencia, Universidad San Sebastián, Santiago, Chile, <sup>9</sup>Departamento de Fisiología, Genética y Microbiología, Universidad de Alicante, Alicante, Spain, <sup>10</sup>Center for Climate and Resilience Research (CR)<sup>2</sup>, Santiago, Chile, <sup>11</sup>Millennium Institute Center for Genome Regulation (CGR), Santiago, Chile

The Cas1 protein is essential for the functioning of CRISPR-Cas adaptive systems. However, despite the high prevalence of CRISPR-Cas systems in thermophilic microorganisms, few studies have investigated the occurrence and diversity of Cas1 across hot spring microbial communities. Phylogenomic analysis of 2,150 Cas1 sequences recovered from 48 metagenomes representing hot springs (42–80°C, pH 6–9) from three continents, revealed similar ecological diversity of Cas1 and 16S rRNA associated with geographic location. Furthermore, phylogenetic analysis of the Cas1 sequences exposed a broad taxonomic distribution in thermophilic bacteria, with new clades of Cas1 homologs branching at the root of the tree or at the root of known clades harboring reference Cas1 types. Additionally, a new family of casposases was identified from hot springs, which further completes the evolutionary landscape of the Cas1 superfamily. This ecological study contributes new Cas1 sequences from known and novel locations worldwide, mainly focusing on under-sampled hot spring microbial mat taxa. Results herein show that circumneutral hot springs are environments harboring high diversity and novelty related to adaptive immunity systems.

## KEYWORDS

Cas1, hot spring, phylogenomic, CRISPR-Cas, casposase

## Introduction

Adaptive immunity in *Bacteria* and *Archaea* is achieved by CRISPR-Cas (clustered regularly interspaced short palindromic repeats and CRISPR-associated genes) systems (Mohanraju et al., 2016; Makarova et al., 2020a). When a foreign nucleic acid invades a prokaryotic cell, it can be recognized by adaptation Cas proteins generating small fragments (spacers) that are stored in a CRISPR array of the host genome separated by the repeats (McGinn and Marraffini, 2019). This information establishes an immunological memory in these cells because spacer-containing transcripts will guide effector Cas to cleave the invasive nucleic acid in future encounters. In most cases, the invading nucleic acid corresponds to viruses (Shmakov et al., 2017a). Four modules, composed of different Cas, have been defined in the functioning of CRISPR-Cas systems (Makarova et al., 2015). The adaptation module integrates the spacers into the host through a complex formed by Cas1 and Cas2, assisted by non-Cas and sometimes other Cas proteins (e.g., Cas4, Cas3, or Cas9) (Koonin and Krupovic, 2015; Amitai and Sorek, 2016; Jackson et al., 2017). Subsequently, the expression module processes multi-spacer transcripts from the CRISPR array (pre-crRNA) to deliver single spacer-containing RNA fragments (crRNAs) that guide Cas of the interference module to act against the invading nucleic acid recognized through complementary bases-pairing with the spacer sequences (Hille et al., 2018). Finally, several proteins or domains in the ancillary/helper module have been described as playing accessory roles (Makarova et al., 2020a). CRISPR-Cas systems are categorized into two classes, six types and over thirty subtypes that could harbor a unique (signature) type gene, and differ in the identity of the associated *cas* genes, mainly those encoding the interference module (Koonin et al., 2017; Shmakov et al., 2017b; Makarova et al., 2020a).

Some CRISPR-Cas systems have been extensively characterized, primarily due to their biotechnological relevance, which has also encouraged the search for new variants in nature (Burstein et al., 2017). However, in the existing databases, some environments are more represented than others, such as human clinical samples versus environmental samples. Among the latter, hot springs are significant for studying CRISPR-Cas because these molecular systems are widespread in the indigenous microorganisms, whether thermophiles or hyperthermophiles (Anderson et al., 2011; Weinberger et al., 2012; Weissman et al., 2019). It has been suggested that temperature impacts the viral diversity and density in these environments by decreasing mutation rates, thereby influencing the occurrence of CRISPR-Cas systems in the host microorganisms (Weinberger et al., 2012; Iranzo et al., 2013; Childs et al., 2014; Westra et al., 2016). Lower mutation rates in thermal environments are explained by the deleterious effect of substitutions at high temperatures (Drake, 2009) which would define a less diverse community than in mesophilic

environments. In thermal environments, the lower virus-prokaryote ratio and lower viral community diversity (Parmar et al., 2018) translate into a lower metabolic cost for the maintenance of the CRISPR-Cas systems against viral infection compared to mesophilic environments (Westra et al., 2016). In this last environment, the spacer catalog has to adapt to more diverse invading nucleic acids (Weinberger et al., 2012; Iranzo et al., 2013; Vale et al., 2015; Westra et al., 2015; Burstein et al., 2016; Van Houte et al., 2016; Broniewski et al., 2020; Meaden et al., 2021). Beyond the high presence of CRISPR-Cas systems in thermophiles, these environments exhibit low microbial complexity compared to mesophilic environments, with fewer microorganisms harboring CRISPR-Cas systems (Burstein et al., 2016; Weissman et al., 2019; Makarova et al., 2020a).

A genetic marker for all CRISPR-Cas systems cannot be established (Koonin and Makarova, 2019; Makarova et al., 2020a). However, the Cas1 protein is the most widespread and evolutionarily conserved *cas* gene (Makarova et al., 2015, 2020a; Koonin and Makarova, 2019) and is essential for CRISPR-Cas adaptive immunity (Krupovic et al., 2014; Amitai and Sorek, 2016; Jackson et al., 2017). Therefore, *cas1* has been used to study the ecology of CRISPR-Cas (Wu et al., 2020). Notably, a protein family composed of Cas1 homologs, called casposases, which is related to the transposition of the carrier mobile genetic element (casposon) has been identified (Krupovic et al., 2014, 2016). The Casposon superfamily has been proposed for the emergence of CRISPR-Cas systems, with their terminal inverted repeats (TIRs) and casposases being the presumed ancestors of CRISPR and CRISPR-associated Cas1, respectively (Koonin and Krupovic, 2015; Krupovic and Koonin, 2016; Mohanraju et al., 2016).

The diversity of Cas1 homologs discovered in new taxa and recently explored environments suggests functions other than those described for immunity (Makarova et al., 2020b), which encourages its study in the natural environment. Despite the high prevalence of CRISPR-Cas systems in hyper/thermophiles, the phylogenomics of the Cas1 protein has not been extensively explored in thermal environments of circumneutral pH. Therefore, to deepen our understanding of the relevance of CRISPR-Cas systems at the community level, the goal of this study was to describe the phylogenetic and environmental diversity of the Cas1 protein in 20 globally distributed hot springs. We hypothesized that these environments harbor new groups of Cas1 homologs not described to date. This study recovered 2,150 Cas1 sequences using 48 metagenomes from 20 hot springs ranging from 42 to 80°C and pH 6 to 9. Our results revealed a correlation between the hot spring dissimilarity observed at Cas1 and taxonomy (16S rRNA), with geographical location as the main explanatory variable of these dissimilarities. Furthermore, several Cas1 homologs did not cluster with reference Cas1 proteins from previously described CRISPR-Cas systems but were positioned at the root of specific

phylogenetic groups in the tree. Finally, some Cas1 from hot springs formed a new family of casposases (proposed family 5).

## Materials and methods

### Study sites, El Tatio sampling, DNA extraction, and sequencing

In this study, we defined a thermophilic temperature range between 40 and 80°C and an approximately neutral pH (6–9) as the most relevant physicochemical parameters to retain hot spring samples (Meyer-Dombard et al., 2005; Inskip et al., 2010; López-López et al., 2013; **Supplementary Figure 1**). Parameters excluded most acidophilic prokaryotes and hyperthermophilic archaea, for which characterization of CRISPR-Cas systems has been previously described (Andersson and Banfield, 2008; Tyson and Banfield, 2008). In total, we analyzed 48 metagenomes, 35 of which were from publicly available data representing sites in North America, South America, and Asia, while 13 were obtained in this study from microbial mats within the El Tatio geyser field, Chile (Figure 1 and **Supplementary Figure 2**). Some El Tatio metagenomes slightly exceeded defined limits (82°C, pH 9.27, **Supplementary Figure 1**), but were retained since their 16S rRNA profiles were similar to other samples from the same location (**Supplementary Figure 3**). The geographic coordinates, physicochemical parameters, DNA source, and accession number of all samples used in this study are available in **Supplementary Table 1**.

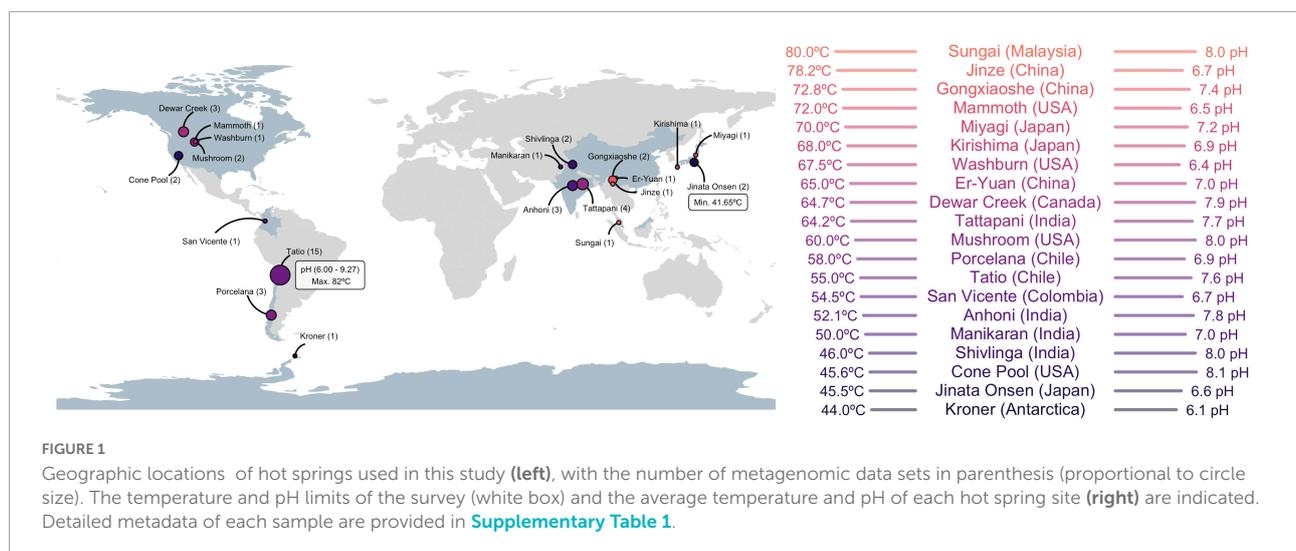
The 13 El Tatio hot springs were selected to cover the entire geothermal field (**Supplementary Figure 2**). Temperatures of the mat samples were measured with a forward-looking infrared camera (Fluke TiS45, WA, USA) and were corroborated on the mat with a multiparameter instrument (WTW multi 340i, NY,

USA). Triplicate samples of approximately 2 ml were collected with a punch from each microbial mat, kept in cryogenic vials containing RNAlater (Thermo Fisher Scientific, Vilnius, Lithuania), and then stored at −80°C until DNA extraction.

DNA extractions of the 13 El Tatio samples were performed according to Alcorta et al. (2018). An equimolar amount of DNA (400 ng) from each replicate was pooled and sent in DNASTable tubes (Biomatrix, San Diego, CA, USA) to the Roy J. Carver Biotechnology Center (University of Illinois at Urbana-Champaign, IL, USA), where libraries were prepared using KAPA HyperPrep (Kapa Biosystems, Roche, Basel, Switzerland) and then sequenced on the Illumina NovaSeq 6000 platform (S1 flowcell, 2 × 150 bp). Quality filtering of reads for the El Tatio samples was performed according to Guajardo-Leiva et al. (2018).

### Metagenome assembly and Cas1 recovery

*De novo* assembly was performed for all 48 metagenomes using SPAdes v3.10.1 (-meta) (Bankevich et al., 2012) except for ERR372908, where MEGAHIT v1.2.9 software (-presets metasensitive) (Li et al., 2015) was used due to memory limitations. Assembly data statistics for all samples used in this study are available in **Supplementary Table 1**. The search for Cas1 orthologs across 48 metagenomes was done with the hmmsearch tool of the HMMER v3.3 package (Eddy, 2011), after ORF prediction with Prodigal v2.6.3 (-p meta) (Hyatt et al., 2010), using the eight updated Cas1 hidden Markov models published by Wu et al. (2020). The *E*-value cut-off (0.01) was set after standardization with one representative sample metagenome (T60, BioSample SAMN15500206 from BioProject PRJNA645256), for which recovered Cas1 candidates were thoroughly curated as described in Moya-Beltrán et al. (2019).



This search yielded 3,556 candidate protein sequences. Putative Cas1 orthologs were compared against acknowledged Cas1 families present in the Conserved Domain Database v.3.16 (CDD; Marchler-Bauer et al., 2017) using CD-search (Marchler-Bauer and Bryant, 2004), hhsearch (Fidler et al., 2016), and RPS-BLAST v2.2.26 (Marchler-Bauer et al., 2002) and were retained if recognized as Cas1 by at least one of the three comparison tools with an *E*-value lower than 0.003 and 0.01 for CD-search and RPS-BLAST, respectively, and a probability higher than 81.6 for hhsearch. Next, a size filter was applied to avoid including possible chimeric proteins and limit the survey to CRISPR-Cas canonical Cas1 (Cas1 representatives smaller than 400 aa (Silas et al., 2017; Wu et al., 2020) were recovered). The 3,414 sequences fulfilling this criterium were then filtered by the relative read abundance in each metagenome. Downstream analyses did not consider sequences with an abundance below 0.1% in each sample. The 2,155 recovered Cas1 sequences were filtered for sequence redundancy at 100% aminoacidic sequence identity to remove sequences absent from public databases, considered here as possibly chimerical. The final data set for this study consisted of 2,150 Cas1 sequences.

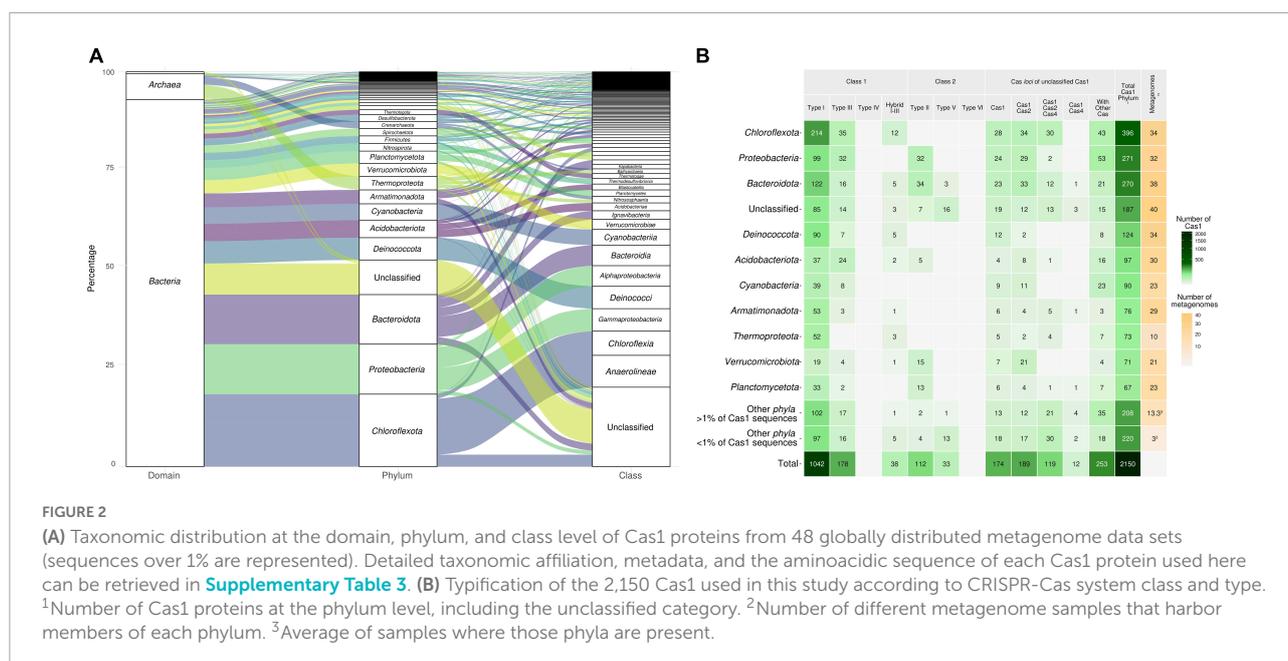
## Taxonomic assignment of Cas1 proteins

We applied three strategies to assign taxonomy to the 2,150 Cas1 sequences recovered in this study. First, we obtained metagenome-assembled genomes (MAGs) of the 48 metagenome data sets according to Alcorta et al. (2020). Taxonomic affiliation of the MAGs was then retrieved using GTDB-tk v0.3.2 software (Chaumeil et al., 2020) with database version R89, identifying *cas1* sequences in contigs housed in

the taxonomically identified MAGs. Second, for *cas1* genes not assigned through MAGs, we used the strategy of Wu et al. (2020) for Cas1 assignment and performed a BLASTp sequence similarity search against the NCBI nr database. Briefly, the five best hits were sorted according to bit score, with the best hit used for assignment if all hits belonged to the same phylum or only if the identity (%) of the best hit was at least 3% higher than the second hit (Wu et al., 2020). Finally, we checked for consistency in the taxonomic affiliation of the Cas1 sequences by retrieving the taxonomic affiliation of the non-Cas gene found in the vicinity of *cas1* in annotated metagenomic contigs (Supplementary Table 2). The assigned taxonomic affiliation of Cas1 sequences is shown in Figure 2A. Aminoacidic sequences and metadata for each sequence are available in Supplementary Table 3.

## Beta-diversity of 16S rRNA and *cas1* genes

For the final set of 2,150 *cas1* genes, reads per Kb per Gb (RPKG) were calculated for each sample using Bowtie2 (Langmead and Salzberg, 2012). Coordinate diversity analyses for the Bray–Curtis dissimilarity index were plotted using non-metric multidimensional scaling (nMDS) in the R package ampvis2 (Andersen et al., 2018). To identify sources of variation considering temperature, pH, altitude, and location (Universal Transverse Mercator, UTM, coordinates), permuted multivariate analysis of variance (PERMANOVA) (Anderson, 2001) was performed with the R package vegan (adonis2, not-sequentially added terms) (Oksanen et al., 2020). The same analyses were performed on the 2,980 16S rRNA genes (over 0.1% RPKG) obtained with MATAM software



(Pericard et al., 2018). Finally, the Mantel test was used to statistically compare the 16S rRNA and Cas1 gene Bray–Curtis matrices. The sequences and taxonomic affiliation of the 16S rRNA genes, as determined with the SILVA 138 SSU database (Quast et al., 2013), are listed in [Supplementary Table 4](#), while rarefaction curves are shown in [Supplementary Figure 4](#).

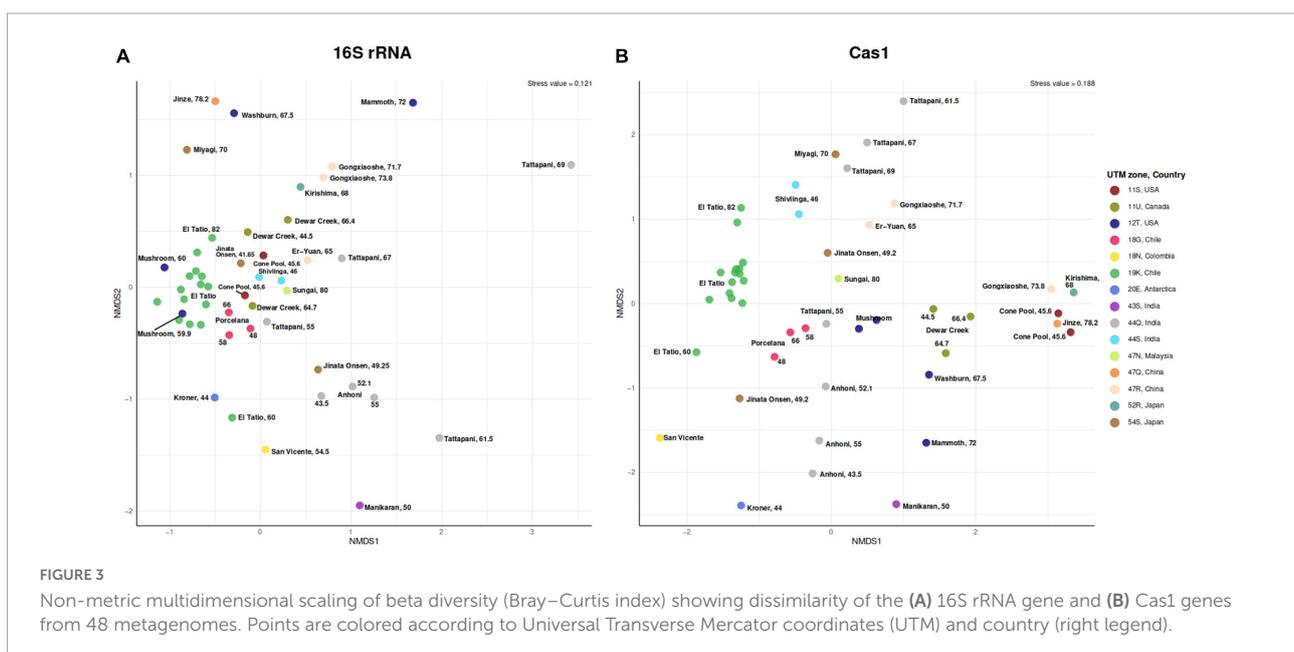
## Cas1 subtypes and phylogenetic analysis

Phylogenetic analyses (ML) were performed with IQtree software (v.1.6.8) (Nguyen et al., 2015) [-m TEST: LG + F + G4 Le and Gascuel model (Le and Gascuel, 2008)] using ultrafast bootstrap (-bb 10,000) (Hoang et al., 2018) after Clustal Omega alignment (Madeira et al., 2019). We followed the recommendation of Wu et al. (2020) to identify Cas1 subtypes using 93 Cas1 reference sequences in the phylogenetic analyses. Additionally, we included caspase genes due to their importance in *cas1* gene evolution (Makarova et al., 2013; Krupovic et al., 2014; Krupovic and Koonin, 2016). The phylogenetic tree was displayed using the iTOL web server (Letunic and Bork, 2019) with the *Streptomyces coelicolor* transposase gene (NP\_626990) as an outgroup to root the tree (Krupovic et al., 2016; Wu et al., 2020). Furthermore, the classification of Cas1 CRISPR-Cas subtypes was analyzed in parallel for each Cas1 using CRISPRCasTyper software (Russel et al., 2020), using mandatory and accessory cas-numerical-score guide typification of the tool. We decided to keep the nomenclature system of CRISPRCasTyper, where complex operons are considered hybrid (six *cas* genes from two or more types with a score of at least six and at least one specific *cas*), and not typified operons are labeled as

ambiguous (non-hybrid operons with two or more *cas* subtypes and the same scoring) or false (neither hybrid nor ambiguous) (Russel et al., 2020).

## Cas1 protein similarity network and gene neighborhood analysis

The Cas1 similarity network analysis was elaborated as described by Cardenas et al. (2016). Briefly, the set of 2,150 Cas1 sequences was clustered using CD-HIT software (Fu et al., 2012) with the parameters outlined for Cas1 clustering (Makarova et al., 2011) (i.e., 90% identity over 75% coverage). The resulting 1,468 representative Cas1 genes were analyzed with BLASTp-all-against-all (default parameters) (Altschul et al., 1990) using an *E*-value of  $10^{-35}$ . Finally, the pairwise bit score was used as the distance for network visualization in Cytoscape 3.9.1 (Shannon et al., 2003) using the *organic* layout. The set of reported contig sequences containing previously identified *cas1* was used to analyze the neighborhood (Moya-Beltrán et al., 2021). Briefly, up to 10 ORFs upstream and/or downstream of *cas1* were recovered and their annotations were retrieved using a GFF file. It should be noted that differences in sequencing quality between the metagenomes used in this study ([Supplementary Table 1](#) and [Supplementary Figure 4](#)) might have affected assembly and therefore the *cas1* genes and vicinity that could be recovered in some samples. Gene products were clustered at a similarity threshold of 0.5 and coverage threshold of 0.33 to obtain representative sequences using MMseqs2 (Steinegger and Söding, 2017). Putative functional assignment of protein clusters was done as described in Moya-Beltrán et al. (2019). Results are summarized in [Supplementary Table 2](#).



## Putative casposase analyses

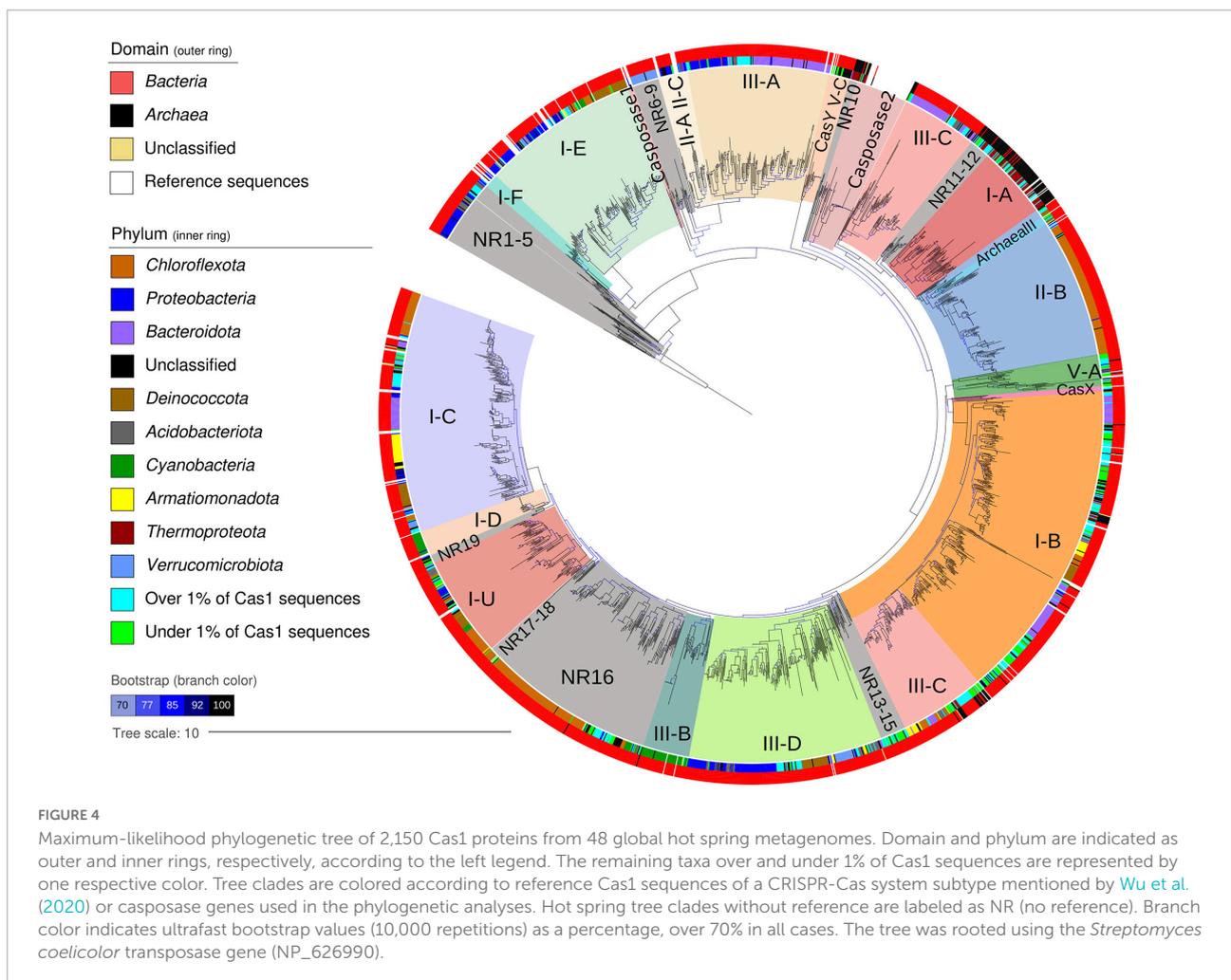
The 174 *cas1* sequences without other *cas*-encoding genes in the vicinity ( $\pm 10$  ORFs, **Figure 2B**) were deemed as putative casposase genes. According to described casposons (Krupovic et al., 2014, 2017), we decided to include *cas1* genes with at least seven genes in the contig and *cas1* not situated at the end of the contig. Seven *cas1* sequences fulfilled this criterion, all of which are allocated with casposase references in the phylogenetic analyses (**Figure 4**); thus, we also included the remaining three Cas1 sequences of the casposase reference clade (**Figure 4**) that present vicinity (Cas1\_1015 was located in the casposase reference clade but without ORFs in the vicinity). In order to determine the casposase family affiliation of these sequences, we reproduced the phylogenetic analyses of Krupovic et al. (2014, 2016). EasyFig v.2.1 (Sullivan et al., 2011) was used to compare contigs of candidate sequences, for which TIRs were searched using TirVish (Gremme et al., 2013). Available TIRs belonging to described casposons (Krupovic et al., 2014) were also used as

queries for identification of TIRs in putative casposase contigs of this study, including the reverse-complement strands.

## Results

### Hot spring sample characterization

Forty-eight metagenomic data sets from 20 hot springs of circumneutral pH, with temperatures in the mesothermophilic to thermophilic range, were recovered from public databases or generated herein (**Supplementary Table 1**) to search for global and local patterns of Cas1 diversity in thermophilic environments. The 48 metagenomic data sets represent nine countries in America and Asia (**Figure 1**). The El Tatio samples submitted here (15 metagenomic data sets), along with those from northern Patagonia (Porcelana, Chile; 3 metagenomic data sets) and Antarctica (Kroner; 1 metagenome), are the only ones from the southern hemisphere within the target physicochemical range (**Figure 1**). El Tatio



sampling encompassed the 10 km<sup>2</sup> area of the high altitude (~4,200 MAMSL) geothermal field (upper, middle, and lower geyser basin, [Supplementary Figure 2](#)) in the Atacama Desert (Chile) ([Glennon and Pfaff, 2003](#)). The 48 samples comprise 36 different temperatures ranging from 41.65°C (Jinata Onsen, Japan) to 82°C (El Tatio, Chile), with an average of 57.9°C and a mode and median of 55°C. An average pH of 7.44 with a mode of 8 and a median of 7.51 was observed. The El Tatio geothermal field recorded the highest (pH 9.72; Pto13) and lowest (pH 6; T82) pH ([Supplementary Table 1](#)). Over 58% of the samples were recovered from microbial mats, while 20% were from sediments and 20% were from water fractions. No information was found on the DNA source for two samples (San Vicente hot spring from Colombia and Mammoth, Liberty Cap Streamers, USA) ([Supplementary Table 1](#)). Extending available hot spring metagenomes to the southern hemisphere balances database samples and brings new taxonomic variants for thermophile microorganisms surveys.

## Frequent hot spring taxa-affiliated Cas1 abundant in thermophilic metagenomes

Of the 2,150 candidates Cas1 finally recovered from 48 metagenomic data sets, 912 (42.4%) were mapped to MAGs (654 MAGs). The remaining 1,238 (57.5%) Cas1 were taxonomically classified *via* BLASTp against the NCBI nr database ([Supplementary Table 3](#)). Approximately 92% of the assignments belonged to the *Bacteria* domain, and 7% to *Archaea*, and were affiliated with 52 phyla and 97 classes. The phylum Chloroflexota accounted for approximately 18% of the total recovered Cas1, represented mainly by classes Chloroflexia (~8%) and Anaerolineae (~6%) ([Figure 2A](#)). The phylum Proteobacteria (~13%) was mainly represented by Alphaproteobacteria and Gammaproteobacteria-affiliated Cas1 sequences. In contrast, the Bacteroidota phylum (~13%) pertained to three main classes (Bacteroidia, Ignavibacteria, and Kapabacteria) ([Figure 2A](#)). Thirty-five phyla and 76 classes were found below the 1% of Cas1 frequency ([Supplementary Table 3](#)). The prevalence of Cas1 in reportedly abundant bacterial hot spring taxa corroborates the importance of CRISPR-Cas systems in hot springs and points out these environments as models to study the environmental microbiology of CRISPR-Cas and their alternative functions ([Mohanraju et al., 2022](#)).

## High frequency of class 1-type I Cas1 proteins in hot springs

Consolidated typification using phylogeny, CRISPRCasTyper, and vicinity approaches ([Supplementary Table 3](#)) allowed us to classify most Cas1 (1,403 of 2,150, 65.2%)

to a CRISPR-Cas type/subtype. An overall predominance of class 1-type I (1,042, 48.4%) was observed ([Figure 2B](#)), whereas class 1-type III was the second most represented type, followed by class 2-type II ([Figure 2B](#)). Thirty-eight Cas1 sequences were classified in type I-type III hybrid operons. No class 1-type IV and class 2-type VI systems were found. For the 174 (8%) Cas1 without *cas* in the neighborhood ([Figure 2B](#)), further analysis was performed to elucidate their relevance as Cas1 not belonging to a *bona fide* CRISPR-Cas locus, as casposase proteins or *cas1*-solo (non-casposase *cas1* without *cas* genes in the neighborhood) (see below). The Cas1 type frequencies observed in hot springs corroborate the predominance of class 1 CRISPR-Cas systems, suggesting that temperature could not be the main driver of CRISPR system types in nature.

## Geographical location as the main driver of *cas1* variance

16S rRNA and *cas1* gene occurrence and abundance *per* metagenome were used to calculate the Bray–Curtis dissimilarity indexes. At the 16S rRNA gene level ([Figure 3A](#)), hot springs belonging to the same UTM zone were more similar than those from different UTM zones. However, samples of intermediate temperature (>55°C < 68°C) from different UTM zones were more similar than samples at the upper and lower end of the temperature range ([Figure 3A](#)). This may indicate the existence of a higher gene flow between hot springs in the middle-temperature range (55–68°C), regardless of their geographic origin. In contrast, temperatures at the upper and lower extremes could restrict gene flow even between nearby geographic zones. Although no significant diversity differences were observed between the *cas1* and 16S rRNA according to the positive correlation (0.57, 0.0001) found by the Mantel test ([Supplementary Table 5](#)), the *cas1* data ([Figure 3B](#)) show that samples were geographically more structured by UTM zone than the results observed for 16S rRNA. The PERMANOVA of Bray–Curtis distances indicates that geographical location explained 47 and 42% of the observed variance in the Cas1 and 16S rRNA, respectively ([Supplementary Table 5](#)). This result is consistent with the nMDS, where the influence of geographic location is more evident at the Cas1 level than at the 16S rRNA level. Other variables, such as temperature, pH, or altitude, scarcely contributed to the variance. Even, a higher sequential effect of geographical location was observed when the sources of variation were added according to the marginal PERMANOVA ([Supplementary Table 5](#)). All these results support the existence of dispersion barriers in hot springs.

To corroborate the relevance of geographical location, we looked for environmental variables affecting the genetic diversity of Cas1. We analyzed Cas1 at the sequence level using a genetic distance matrix (percent identity). The marginal PERMANOVA with distance matrix values revealed that all

sources of variation considered here poorly explained the differences at the sequence level (Supplementary Table 5), suggesting that Cas1 critical structure is very conserved in hot springs, at least in the canonical protein size used here.

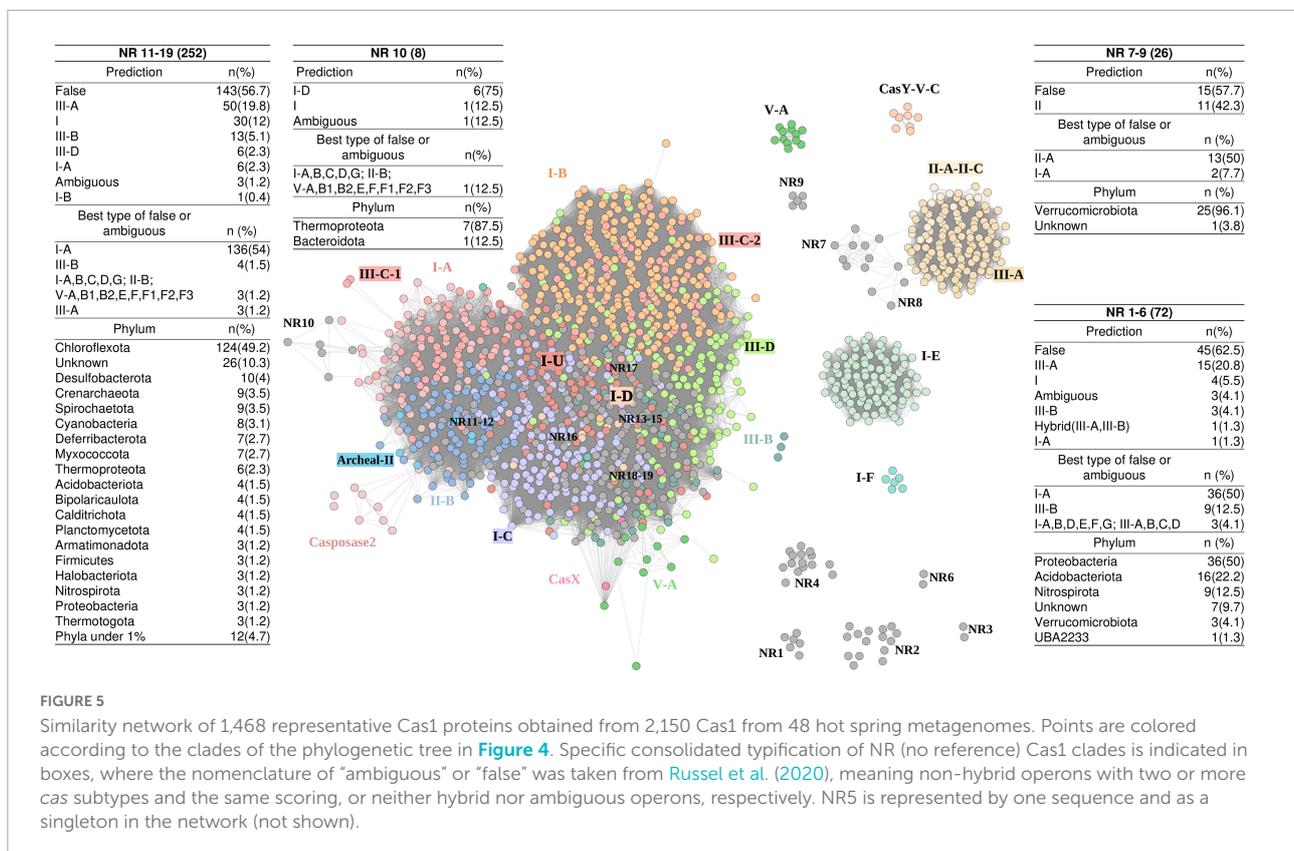
## Novel subclades of hot spring Cas1

A rooted phylogenetic tree was constructed with the 2,150 sequences of Cas1 proteins from the 48 globally distributed hot springs (Figure 4), along with 93 Cas1 reference sequences that helped typification (Wu et al., 2020). Most hot spring Cas1 sequences clustered with the reference sequences, showing that hot spring Cas1 are linked to CRISPR-Cas systems of diverse known subtypes (Figure 4). However, 19 clades without reference sequences were also obtained, harboring poorly classified or unclassified Cas1 sequences (labeled NR, i.e., No Reference). Except for subclade NR16, NR clades were usually positioned next to the root of the tree or the root of internal reference Cas1 clades (Figure 4, NR marked in gray). NR clades and their position indicate that some Cas1 from hot springs are infrequent in databases or completely new, and also suggest that Cas1 sequences close to casposase clades may be related to casposons.

Overlaid taxonomic affiliation showed that 52 microbial phyla were represented at least once in the tree (Figure 4,

outer and inner rings). Cas1 from the same bacterial phyla frequently clustered together, and below the phylum level, no clear associations were recovered (data not shown). However, Cas1 proteins from predominant phyla such as *Chloroflexota* (396 sequences) and *Proteobacteria* (271 sequences) were distributed in several clades within the tree (Figure 4). Most of the 90 Cas1 sequences from members of the phylum *Cyanobacteria* were clustered into clades recognized as I-D and III-B (Figure 4). Metadata overlay on the tree (e.g., temperature) was of poor value for revealing specific data patterns (Supplementary Figure 5), except for Cas1 clades I-A and I-B belonging to *Archaea*, which were associated with slightly warmer temperatures. Regarding the taxonomy of NR subclades (358 total sequences), most belonged to *Chloroflexota* (124, 35%) and *Proteobacteria* (39, 11%), whereas taxonomically unassigned (phylum-level) Cas1 sequences ranked third (34, 9.5%), which could suggest new Cas1 homologs, especially for those deep clades. In general, phyla that harbored most of the Cas1 sequences (Figure 2) also harbored most of the NR members (over 2%, Figure 5), indicating that predominant hot spring taxa could harbor rare Cas1 homologs.

To further explore the nature and characteristics of the Cas1 NR subclades, we constructed a similarity network using the 1,468 representative Cas1 proteins of the aforementioned phylogenetic tree (Figure 4). Each node of the network (Figure 5) represents a Cas1 protein, and edges correspond to



the bit score of an all-versus-all BLASTp analysis. The largest network module contains three principal regions conforming a core, where (1) most I-A, II-B, III-C-1, and Archaeal-II Cas1 sequences are separated from the region composed of (2) I-C, and (3) I-B, III-C-2, and III-D subtype proteins. Some nodes are separated by long edges and arranged as “satellites”: Casposase2, CasX, and some Cas1 of subclade V-A (Figure 5), which indicates unusual Cas1 homologs. The second-largest module harbors Cas1 of subtypes II-A-II-C and III-A, together with “satellite” sequences (NR7 and NR8, Figure 5). The third module is composed exclusively of I-E sequences, which in Figure 4 also clustered apart with I-F representatives (a minor yet separate module in the network; Figure 5). Finally, Cas1 of the rare subtype CasY formed an isolated module. Several small NR modules are isolated from the rest (NR1–NR6). However, the biggest NR clade of Figure 4 (NR16) was inside the main network module (Figure 5), suggesting new varieties of Cas1 similar to traditional Cas1 sequences.

Most NR Cas1 clades of the phylogenetic tree were classified as false or ambiguous (210, 9.7%), but also several sequences were affiliated with a CRISPR-Cas system type/subtype (Figure 5 and Supplementary Table 3). According to the network arrangement, the main module harbors 252 sequences of subclades NR11–NR19, where 146 (58%) are classified as false or ambiguous, and 106 (42%) effectively belong to a known CRISPR-Cas system (Figure 5 and Supplementary Table 6). Interestingly, modules NR7, NR8, and NR9 only include sequences classified as type II Cas1 or false. The NR10 cluster is predominantly of type I-D Cas1 sequences and is displayed as “satellite” of the main cluster, which is also observed for NR7, NR8, and Cas1 sequences of the Casposase2 clade (Figure 5). These results show that Cas1 from hot springs harbor rare sequence variants. Deeper analyses of “satellite” casposase clade allowed us to identify new Cas1 unrelated to CRISPR-Cas immunity.

## New casposase genes from hot springs

Performing the casposase phylogenetic analyses of Krupovic et al., 2014, 2016, 6 of 10 Cas1-solo from hot springs were located in a monophyletic clade in the outgroup, close to Cas1 subtype I-A (Figure 6), representing a novel branch of casposases. We propose that this clade corresponds to a new family of casposases (family 5) discovered in hot springs. Other hot spring casposases grouped with already known families (Figure 6). New family 5 and Cas1 proteins belonging to known casposase families show great neighborhood genetic diversity (Figure 7 and Supplementary Figure 6). Identity comparison also revealed great variation, where only contigs harboring Cas1\_938 (Miyagi, Japan) and Cas1\_1200 (Washburn, USA) presented greater similarity (Figure 7). The same was observed for the gene content of the contig, where the most distant

contig (Cas1\_1244) is the only one harboring DNAPol B. TIRs were identified in five contigs (Cas1\_18, Cas1\_938, Cas1\_1200, Cas1\_1226, and Cas1\_1269). For sequence Cas1\_1226, the TIR was identified by alignment with the TIR *Candidatus* “Acetothermum autotrophicum” (Krupovic et al., 2014). These results reveal the remarkable variation of genetic content inside family 5; however, casposases of this family show a high identity between them compared to other families (Supplementary Figure 7). The genetic neighborhood of remaining non-family 5 putative casposase gene contigs are available in Supplementary Figure 6.

## Discussion

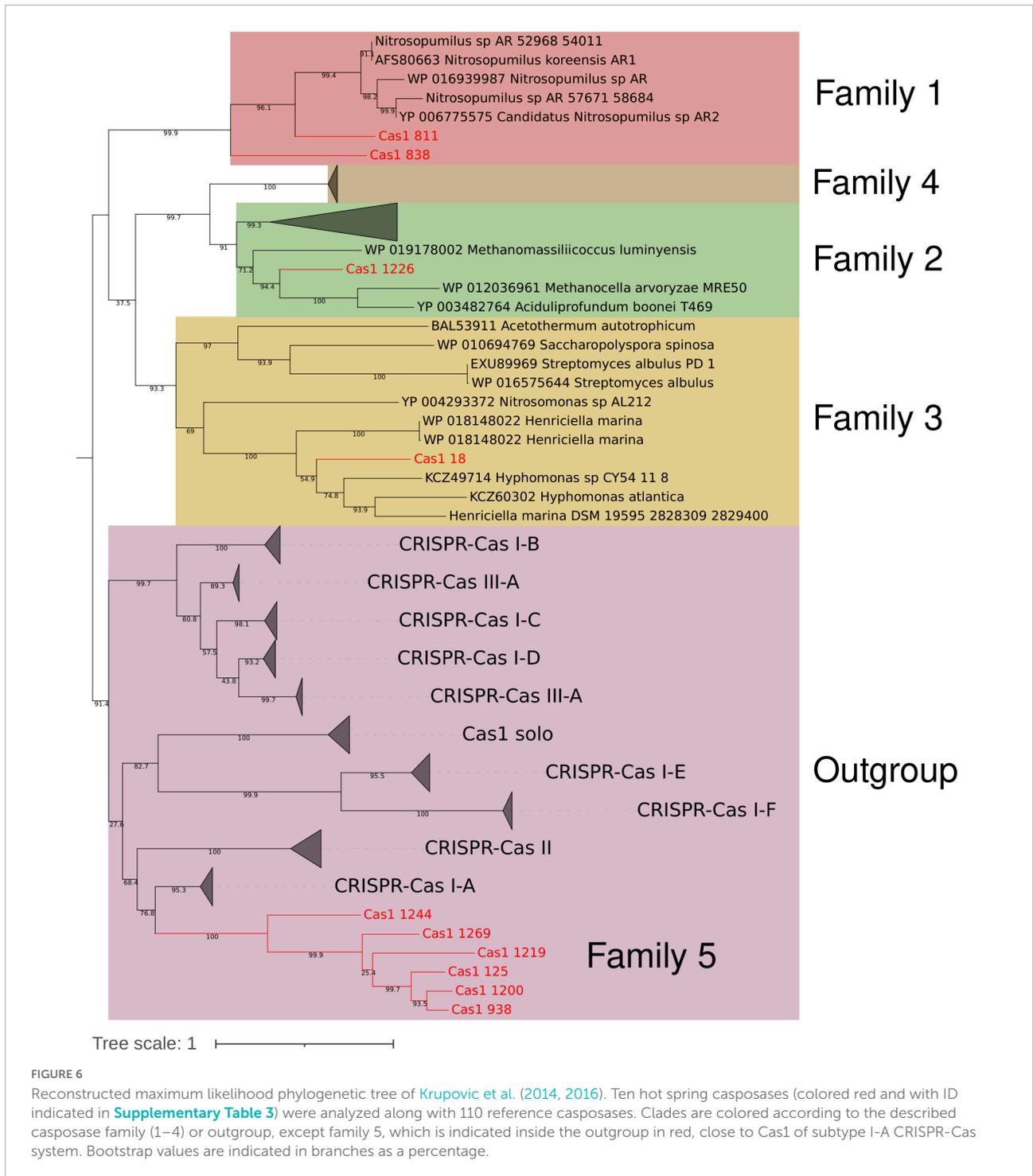
### Expanding the *cas1* gene information to the southern hemisphere

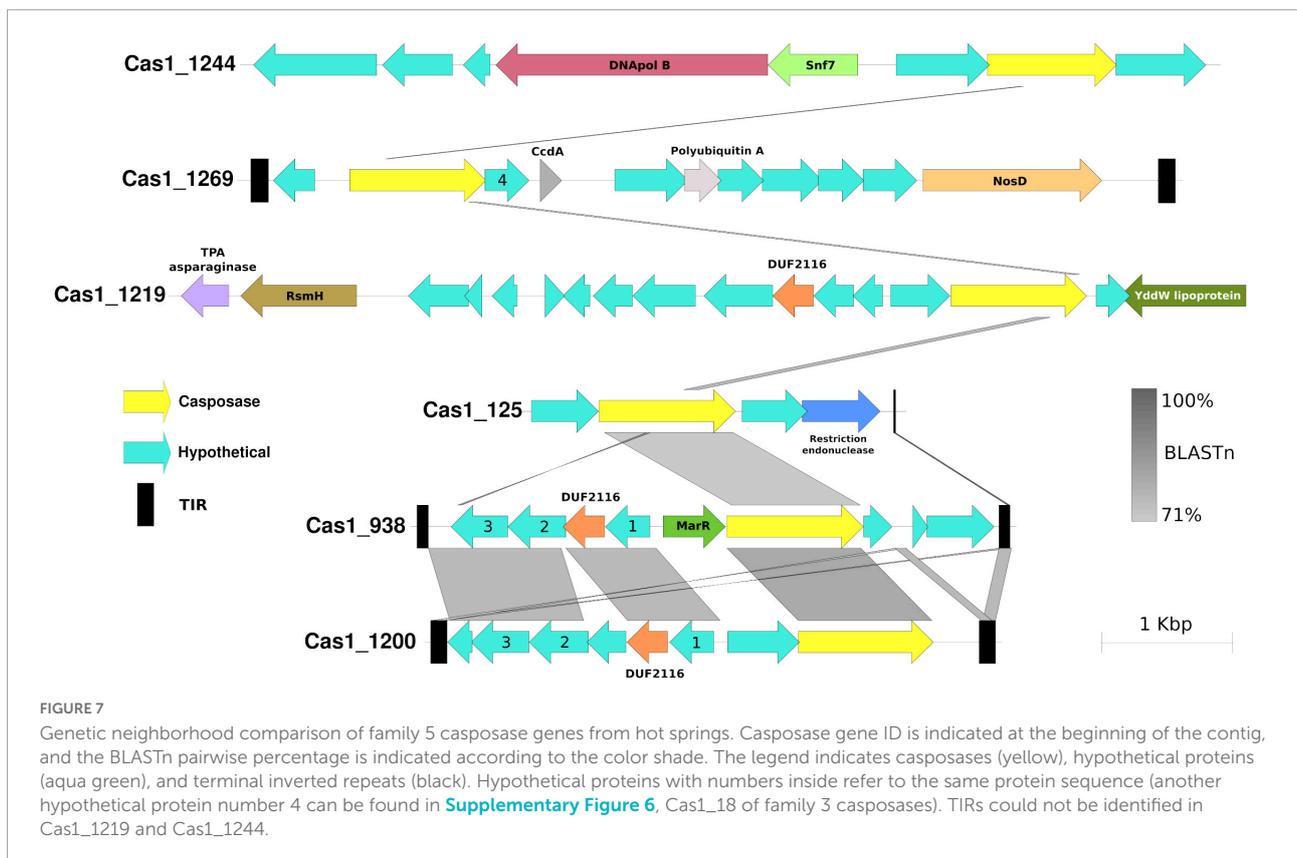
The metagenomes used in this study included new El Tatio metagenomes from El Tatio geothermal field in Chile, expanding the metagenomic data available for hot springs from underrepresented geographical zones. In addition, this study contributed to the expansion of the known Cas1 sequence from the southern hemisphere, including Antarctica (Figure 1). The global and regional distribution of the metagenomic data used allowed us to compare the diversity of 16S rRNA and Cas1 on a continental scale, but also locally, as in the case of the El Tatio geyser field. The ranges of temperature (41–80°C) and pH (approximately 6–8) used for this study have been defined as suitable for thermophilic microorganisms (Zablocki et al., 2018; Merino et al., 2019) and reveal that metagenomic samples in the mesothermophilic range of temperature (approximately 55–65°C) are globally similar, corroborating that temperature and pH are the main drivers of microbial diversity in hot springs (Massello et al., 2020). Nevertheless, the local adaptation revealed by some Cas1 suggests the presence of dispersal barriers to gene flow which may be associated with variables not quantified in this study (Figure 3B).

The metagenomic data sets used here corroborate the presence of previously described phyla abundant in hot springs (Klatt et al., 2011, 2013; Inskeep et al., 2013; López-López et al., 2013; Bolhuis et al., 2014; Sharp et al., 2014; Strazzulli et al., 2017). Most hot spring metagenomes share a small core of highly represented taxa, such as the phylum *Chloroflexota*, and several minor phyla (Figure 2 and Supplementary Figure 3). However, the species turnover evidenced by the 16S rRNA gene Bray–Curtis dissimilarity index, was less than the turnover for the *cas1* gene (Figure 3), which was more geographically structured. This is consistent with the degree of isolation and local evolution previously suggested for thermal environments and some phyla, such as thermophilic *Cyanobacteria* (Finsinger et al., 2008; Kunin et al., 2008; Ionescu et al., 2010). There are some exceptions, such as the case of Tattapani, which showed high

dissimilarity for the 16S rRNA gene, but not for *cas1* (Figure 3). One possible explanation could be a local adaptation of CRISPR-Cas systems to similar conditions and, for example, confronting a similar viral community. Meanwhile, other samples, such as from Gongxiaoshe, showed high similarity in the 16S rRNA gene but great dissimilarity with *cas1* (Figure 3). This could be due to the presence of similar bacterial communities

versus locally different viral communities, which would support differential adaptation of the host adaptation module revealed by *cas1*. We hypothesize that Cas1 may expose local adaptations due to the specificity of virus-host relationships (Zablocki et al., 2018) mediated by CRISPR-Cas systems in hot springs. This hypothesis is consistent with the argued rapid evolution observed in extreme environments (Li et al., 2014) and the





specificity of prokaryotic genes (Coelho et al., 2022), which could also be necessary for thermophilic viral communities. *Cas1* beta-diversity variations can also suggest that the same host could have existed at the beginning of several hot spring communities, where today viruses may help reveal evolutionary changes in these host communities.

## High *cas1* abundance in prevalent hot spring phyla

Abundances of *cas1* and 16S rRNA genes showed similarities in each sample regardless of hot spring temperature ([Supplementary Figure 3](#)), especially for predominant phyla such as *Chloroflexota*, *Proteobacteria*, *Cyanobacteria*, and *Bacteroidota*, which highlights the role of CRISPR-Cas systems. *Cas1* protein diversity analyses at the community level are scarce, and few surveys consider predominant hot springs taxa and CRISPR-Cas types, particularly for *Chloroflexota* and *Bacteroidota* phyla. Wu et al. (2020) observed that *cas1* prevalence in soil samples increased with temperature in some taxa (including *Chloroflexota* and *Deltaproteobacteria*) while *rplB* abundance decreased. Differences here could be explained due to the stability of hot springs versus temperature as altering the microbiostasis of soil. However, it should be noted that the low sequencing depth of some

samples used here (M46\_SRR2625865, M46\_SRR2626160, M61\_SRR3961741, M50\_ERR1543536, M4564\_SRR6941191, and M65\_MR4530144) could overestimate diversities found. In any case, the maintenance or removal of low-coverage samples does not alter the statistical significance of the positive correlation between 16S rRNA and *cas1* genes ([Supplementary Table 5](#)). As suggested (Weinberger et al., 2012; Irazzo et al., 2013), temperature changes could determine differential fitness for CRISPR-Cas systems (affecting the viral diversity/density of specific taxa) in the long term, which may explain the correlation observed in soil. A high prevalence of *Cas1* in major taxa from hot springs can reveal a community where predominant species, mainly phototrophic, could be competition and adaptive defense specialists due to low viral diversity/density. CRISPR-Cas in predominant taxa from hot springs, especially in phototrophic species (*Chloroflexota* and *Cyanobacteria* phyla), could maintain the steady state of the system, ensuring the inflow of energy.

## The novelty of the *cas1* gene in hot springs

Reference sequence-guided typing (Wu et al., 2020) of previously described *Cas1* sequences helped identify the CRISPR-Cas system category likely to be most relevant in these

thermal environments. However, this approach has limitations because the *in silico* classification of CRISPR-Cas into a specific type/subtype must be based not only on phylogeny but also on the sequence similarity, genetic vicinity, domains, and catalytic residues (Makarova et al., 2020a). For this reason, we further employed CRISPRCasTyper software (Russel et al., 2020), to retrieve information in the contigs harboring *cas1* that showed type I and III CRISPR-Cas as predominant in hot springs (Figure 2), as well as in databases (Koonin et al., 2017; Crawley et al., 2018). Type I and III systems accounted for 48 and 8.2%, respectively, of the Cas1 obtained herein, which is below the total distribution of CRISPR-Cas systems reported in nature (Crawley et al., 2018). Despite type IV systems having been previously found in hyper/thermophiles (Jung et al., 2016; Taylor et al., 2019; Makarova et al., 2020a), we did not detect type IV Cas1 genes in the analyzed metagenomes. This might be due to the almost total absence described for the adaptation module in type IV CRISPR-Cas systems (Pinilla-Redondo et al., 2020), being overlooked here due to the methodological approach used. Conversely, the fact that class 1 systems are predominant in both thermophilic and mesophilic environments suggests that selective pressure is not related to temperature and mutation rates, but to specific mechanistic properties (Shmakov et al., 2017b). In this sense, class 1 systems could be more versatile in performing several functions than class 2 systems or in escaping anti-immunity mechanisms, for example, evolving just one subunit targeted by anti-CRISPR. Regarding class 2, 146 (6.7%) of Cas1 were classified as type II and V (Figure 2). Most of the Cas1 classified as type V belong to rare phyla such as *Patescibacteria* or unclassified bacteria (Figure 2), which encourages the search for new variants of this type already described in thermophiles (Chen et al., 2019; Tian et al., 2020). The absence of type VI Cas1 could be expected because these CRISPR-Cas systems target RNA (Makarova et al., 2020a), maybe pointing to scarce foreign RNA entering the cell. However, as mentioned for system IV, *cas1* genes associated with type IV systems have only been described in a few subtypes (Makarova et al., 2020a). Furthermore, described bacterial species that harbor type VI systems are mesophilic and related to humans or pets, which suggests a low prevalence of this type in hot springs.

Several Cas1 sequences could not be classified to a described system type/subtype despite being together with other *cas* genes (Figure 2). Given the genetic context, phylogeny, and network data, we speculate that some of these Cas are part of novel CRISPR-Cas systems. It should be noted that our approach was based on the tree topology, *cas* locus, and CRISPR array [using CRISPRCasTyper (Russel et al., 2020)]; however, missing contig information could hinder its assignment to a CRISPR-Cas system. Nevertheless, Cas2 and Cas4 were the most frequent *cas* gene in NR Cas1 clades. Regarding non-*cas* genes, prevalent genes were encoding hypothetical proteins as not in the CDD database (426 ORFs), RNase H-like (26 ORFs), and DUF697 or

DUF370 domain-containing proteins. The fact that most gene neighborhoods are related to DNA metabolism corroborates the evidence of non-canonical functions of CRISPR-Cas systems and Cas1 (Sampson and Weiss, 2013; Krishnan et al., 2020; Mohanraju et al., 2022). Ongoing work will allow us to eventually characterize new molecular systems involving Cas proteins, which exceeds the current objectives of this work.

Our phylogenetic reconstruction is in agreement with the topology retrieved by Wu et al. (2020), but also includes several clades without reference sequences (NR), located at the root of the tree and in several internal clades (Figure 4). The phylum *Chloroflexota* harbors most of the Cas1 NR clades, which may be explained by the dominance of diverse members of this phylum in hot springs. The *Chloroflexota* Cas1 sequences were distributed in tree clades II-B, NR16, NR18, and I-U (Figure 4 and Supplementary Table 6), suggesting great diversity of adaptive immunity in this phylum. According to the data (Makarova et al., 2011, 2013, 2020a; Burstein et al., 2016), hot springs Cas1 from *Chloroflexota* includes types I and III, with a predominance of type I-A. However, in most cases, members that group in the tree with type II-B reference sequences (WP080019870 and WP011139432) belong to I-B (Supplementary Table 3), suggesting horizontal transfer of Cas1. The expected scenario was observed for *Cyanobacteria*, where sequences were located as described in I-E (Makarova et al., 2013; Burstein et al., 2016), III-B (Makarova et al., 2013, 2020a; Burstein et al., 2016), and the almost phylum-exclusive I-D (Cai et al., 2013). Our results maintain the absence of type II for *Cyanobacteria* and *Chloroflexota* (Makarova et al., 2020a; Figure 4), with the majority of Cas1 of this type found in *Proteobacteria* and *Bacteroidota*, more represented by non-photosynthetic taxa (Figure 2B and Supplementary Table 3), suggesting a relationship between photosynthesis and the scarcity of the Cas9 protein. Overall, it is difficult to describe the diversity of Cas1 inside these hot spring-predominant taxa using taxonomy (despite the exclusiveness of some subtypes, e.g., I-D). As mentioned for the CRISPR-Cas system (Makarova et al., 2015, 2020a), the presence of the same type/subtype in several taxa could suggest no specific function of Cas1 in a host. Furthermore, the fact that hot springs did not reveal a new association regarding phylum-Cas1 type descriptions suggests that temperature has a minor selective effect for the CRISPR-Cas type/subtype. High horizontal gene transfer and viral infection events regulating CRISPR-Cas systems (Makarova et al., 2015; Koonin et al., 2017; Landsberger et al., 2018) could also blur the evolutionary history of Cas1 in hot springs, a hypothesis framed in the “guns for hire” model (Koonin et al., 2020). However, the case of I-D in *Cyanobacteria* points to a particular virus infecting them or specific function not necessarily related to adaptive immunity (Mohanraju et al., 2022), which could maintain Cas1 as a “not for hire” gun.

Finally, Casposase analysis highlights the novelty of Cas1 from hot springs and confirms the relevance of CRISPR-Cas in these environments. Previous work on hot springs Cas1 diversity defined four casposase gene families and a Cas1-solo outgroup, indicating that those Cas1-solo were probably vestigial genes due to non-conserved catalytic residues (Krupovic et al., 2014, 2016). With new metagenomic data sets available today, we have identified a new casposase gene group (proposed family 5, Figure 6), the closest in the tree to the Cas1-solo outgroup mentioned by Krupovic et al. (2014, 2016) and next to subtype I-A Cas1. The CRISPR-Cas subtype I-A has a majority representation in the *Archaea* domain (Makarova et al., 2015), which is observed in Cas1 subtype I-A of the tree (Figure 4), but also for family 5 casposases, suggesting a vestigial metabolic function of Cas1 related with that domain. Conversely, family 5 shows conserved catalytic site residues (Supplementary Figure 7), suggesting that they could be active enzymes. Nevertheless, casposase Cas1\_1244 is the only family member with DNAPol B and shows the lowest identity value with the rest of the family 5 (Supplementary Figure 7), whose remaining members share over 50% identity. Active site residues and the absence of DNAPol in most of the family 5, contrary to other Casposase families (Krupovic et al., 2014, 2017), suggest that they are functioning as non-self-replicative transposons, which is in line with the non-relationship of casposons with eukaryotic self-synthesizing transposons (Krupovic et al., 2014). We speculate that family 5 represents an intermediate stage between casposases from families 1 to 4 and Cas1 of CRISPR-Cas systems, suggesting that inactive Cas1-solo [group 1 of Krupovic et al. (2014)] and family 5 might represent recent ancestors in the evolution of Cas1. The HTH C-terminal domain identified in family 2 casposase (Hickman and Dyda, 2015) was not observed in family 5 casposase (data not shown), supporting its position as an ancestor of CRISPR-Cas Cas1. The high diversity of the genetic context of family 5 casposons, composed of poorly conserved hypothetical proteins (Figure 7), as well as their low sequence similarity concerning other families (Supplementary Figure 7), also suggest their rapid evolution, perhaps influenced by several potential horizontal gene transfer events. Only *loci* Cas1\_938 and Cas1\_1200 of family 5 share more than one hypothetical protein, which is intriguing because of their origin from very distant hot springs (Japan and USA, respectively, Supplementary Tables 1, 3). New casposases help to shed light on the function and evolution of hot spring casposons which will also contribute to the study of the evolution of CRISPR-Cas systems, thereby revealing potential new features that would allow for better elucidation of the origin of the system.

The present study extends the knowledge of Cas1 diversity in thermal environments, where ecological diversity was associated with local characteristics according to geographical origin. Phylogeny and network analyses reveal new Cas1

homologs, including a new family of casposons that formally extends the currently known diversity of the gene. This work could contribute to a better understanding of the evolution of CRISPR-Cas systems by describing new variants in new genetic contexts obtained from new hot springs metagenomes. This study also corroborates that hot springs are suitable environments for obtaining novel information on CRISPR-Cas ecology and evolution, and could contribute to understanding the higher prevalence of CRISPR-Cas systems in these environments.

## Data availability statement

All data generated or analyzed during this study are included in this published article and **Supplementary material**. Metagenomic datasets from El Tatio can be found in the NCBI database under BioProject PRJNA858297 (<https://www.ncbi.nlm.nih.gov/sra/>).

## Author contributions

OS, SG-L, CB, CR, and BD made the field sampling. OS, SG-L, AM-B, and JT-L made *in silico* analyses. RQ, FJMM, and BD contributed significantly to the research design and writing process. OS and BD conceived the study and wrote the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This work was financed in part by FONDECYT regular N° 1190998 (ANID) and Iniciativa de Investigación UnACH 2021-157-Unach. OS and JT-L were supported in part by ANID National Doctoral Scholarship (Beca de Doctorado Nacional ANID) N° 21172022 and 21171048, respectively. SG-L was supported by ANID FONDECYT Postdoctoral N° 3210547. AM-B and RQ were supported by Centro Ciencia and Vida, FB210008, Financiamiento Basal para Centros Científicos y Tecnológicos de Excelencia de ANID, and FONDECYT regular N° 1221035 (ANID). FJMM acknowledged research support by the Conselleria d'Innovació, Universitat, Ciència i Societat Digital from Generalitat Valenciana, research project PROMETEO/2021/057. BD acknowledged the Millennium Institute Center for Genome Regulation, Project ICN2021-044 supported by the ANID Millennium Scientific Initiative (Chile).

## Acknowledgments

OS thanks Pablo Vergara for his technical help. BD thanks El Tatio local communities Toconce and Caspana.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.1069452/full#supplementary-material>

## References

- Alcorta, J., Alarcón-Schumacher, T., Salgado, O., and Díez, B. (2020). Taxonomic novelty and Distinctive genomic features of hot spring cyanobacteria. *Front. Genet.* 11:568223. doi: 10.3389/fgene.2020.568223
- Alcorta, J., Espinoza, S., Viver, T., Alcamán-Arias, M. E., Trefault, N., Rosselló-Móra, R., et al. (2018). Temperature modulates *Fischerella thermalis* ecotypes in Porcelana hot spring. *Syst. Appl. Microbiol.* 41, 531–543. doi: 10.1016/j.syapm.2018.05.006
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Amitai, G., and Sorek, R. (2016). CRISPR-Cas adaptation: Insights into the mechanism of action. *Nat. Rev. Microbiol.* 14, 67–76. doi: 10.1038/nrmicro.2015.14
- Andersen, K. S., Kirkegaard, R. H., Karst, S. M., and Albertsen, M. (2018). ampvis2: An R package to analyse and visualise 16S rRNA amplicon data. *bioRxiv* [Preprint]. doi: 10.1101/299537
- Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 26, 32–46. doi: 10.1111/j.1442-9993.2001.01070.pp.x
- Anderson, R. E., Brazelton, W. J., and Baross, J. A. (2011). Using CRISPRs as ametagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage. *FEMS Microbiol. Ecol.* 77, 120–133. doi: 10.1111/j.1574-6941.2011.01090.x
- Andersson, A. F., and Banfield, J. F. (2008). Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* 320, 1047–1050. doi: 10.1126/science.1157358
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Bolhuis, H., Cretoiu, M. S., and Stal, L. J. (2014). Molecular ecology of microbial mats. *FEMS Microbiol. Ecol.* 90, 335–350. doi: 10.1111/1574-6941.12408
- Broniewski, J. M., Meaden, S., Paterson, S., Buckling, A., and Westra, E. R. (2020). The effect of phage genetic diversity on bacterial resistance evolution. *ISME J.* 14, 828–836. doi: 10.1038/s41396-019-0577-7
- Burstein, D., Harrington, L. B., Strutt, S. C., Probst, A. J., Anantharaman, K., Thomas, B. C., et al. (2017). New CRISPR-Cas systems from uncultivated microbes. *Nature* 542, 237–241. doi: 10.1038/nature21059
- Burstein, D., Sun, C. L., Brown, C. T., Sharon, I., Anantharaman, K., Probst, A. J., et al. (2016). Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nat. Commun.* 7:10613. doi: 10.1038/ncomms10613
- Cai, F., Axen, S. D., and Kerfeld, C. A. (2013). Evidence for the widespread distribution of CRISPR-Cas system in the Phylum Cyanobacteria. *RNA Biol.* 10, 687–693. doi: 10.4161/rna.24571
- Cardenas, J. P., Quatrini, R., and Holmes, D. S. (2016). Aerobic lineage of the oxidative stress response protein rubrerythrin emerged in an ancient microaerobic, (hyper)thermophilic environment. *Front. Microbiol.* 7:1822. doi: 10.3389/fmicb.2016.01822
- Chaumeil, P. A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2020). GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36, 1925–1927. doi: 10.1093/bioinformatics/btz848
- Chen, L. X., Al-Shayeb, B., Méheust, R., Li, W. J., Doudna, J. A., and Banfield, J. F. (2019). Candidate phyla radiation roizmanbacteria from hot springs have novel and unexpectedly abundant CRISPR-cas systems. *Front. Microbiol.* 10:928. doi: 10.3389/fmicb.2019.00928
- Childs, L. M., England, W. E., Young, M. J., Weitz, J. S., and Whitaker, R. J. (2014). CRISPR-induced distributed immunity in microbial populations. *PLoS One* 9:e101710. doi: 10.1371/journal.pone.0101710
- Coelho, L. P., Alves, R., del Río, Á.R., Myers, P. N., Cantalapiedra, C. P., Giner-Lamia, J., et al. (2022). Towards the biogeography of prokaryotic genes. *Nature* 601, 252–256. doi: 10.1038/s41586-021-04233-4
- Crawley, A. B., Henriksen, J. R., and Barrangou, R. (2018). CRISPRdisco: An automated pipeline for the discovery and analysis of CRISPR-Cas systems. *Cris. J.* 1, 171–181. doi: 10.1089/crispr.2017.0022
- Drake, J. W. (2009). Avoiding dangerous missense: Thermophiles display especially low mutation rates. *PLoS Genet.* 5:e1000520. doi: 10.1371/journal.pgen.1000520
- Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* 7:1002195. doi: 10.1371/journal.pcbi.1002195
- Fidler, D. R., Murphy, S. E., Courtis, K., Antonoudiou, P., El-Tohamy, R., Ient, J., et al. (2016). Using HHsearch to tackle proteins of unknown function: A pilot study with PH domains. *Traffic* 17, 1214–1226. doi: 10.1111/tra.12432
- Finsinger, K., Scholz, I., Serrano, A., Morales, S., Uribe-Lorio, L., Mora, M., et al. (2008). Characterization of true-branching cyanobacteria from geothermal sites and hot springs of Costa Rica. *Environ. Microbiol.* 10, 460–473. doi: 10.1111/j.1462-2920.2007.01467.x
- Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565
- Glennon, J. A., and Pfaff, R. M. (2003). The extraordinary thermal activity of El tatio geyser field, antofagasta region, Chile. *GOSA Trans.* 8, 31–78.
- Gremme, G., Steinbiss, S., and Kurtz, S. (2013). GenomeTools: A comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 10, 645–656. doi: 10.1109/TCBB.2013.68
- Guajardo-Leiva, S., Pedrós-Alió, C., Salgado, O., Pinto, F., and Díez, B. (2018). Active crossfire between cyanobacteria and cyanophages in phototrophic mat communities within hot springs. *Front. Microbiol.* 9:2039. doi: 10.3389/fmicb.2018.02039
- Hickman, A. B., and Dyda, F. (2015). The casposon-encoded Cas1 protein from *Aciduliprofundum boonei* is a DNA integrase that generates target site duplications. *Nucleic Acids Res.* 43, 10576–10587. doi: 10.1093/nar/gkv1180
- Hille, F., Richter, H., Wong, S. P., Bratovič, M., Ressel, S., and Charpentier, E. (2018). The biology of CRISPR-Cas: Backward and Forward. *Cell* 172, 1239–1259. doi: 10.1016/j.cell.2017.11.032
- Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281

- Hyatt, D., Chen, G. L., LoCasio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Inskeep, W. P., Jay, Z. J., Tringe, S. G., Herrgård, M. J., and Rusch, D. B. (2013). The YNP metagenome project: Environmental parameters responsible for microbial distribution in the yellowstone geothermal ecosystem. *Front. Microbiol.* 4:67. doi: 10.3389/fmicb.2013.00067
- Inskeep, W. P., Rusch, D. B., Jay, Z. J., Herrgård, M. J., Kozubal, M. A., Richardson, T. H., et al. (2010). Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. *PLoS One* 5:e9773. doi: 10.1371/journal.pone.0009773
- Ionescu, D., Hindiyyeh, M., Malkawi, H., and Oren, A. (2010). Biogeography of thermophilic cyanobacteria: Insights from the Zerka Ma'in hot springs (Jordan). *FEMS Microbiol. Ecol.* 72, 103–113. doi: 10.1111/j.1574-6941.2010.00835.x
- Iranzo, J., Lobkovsky, A. E., Wolf, Y. I., and Koonin, E. V. (2013). Evolutionary dynamics of the prokaryotic adaptive immunity system CRISPR-Cas in an explicit ecological context. *J. Bacteriol.* 195, 3834–3844. doi: 10.1128/JB.00412-13
- Jackson, S. A., McKenzie, R. E., Fagerlund, R. D., Kieper, S. N., Fineran, P. C., and Brouns, S. J. J. (2017). CRISPR-Cas: Adapting to change. *Science* 356:eaa15056. doi: 10.1126/science.aal5056
- Jung, T. Y., Park, K. H., An, Y., Schulga, A., Deyev, S., Jung, J. H., et al. (2016). Structural features of Cas2 from *Thermococcus onnurineus* in CRISPR-cas system type IV. *Protein Sci.* 25, 1890–1897. doi: 10.1002/pro.2981
- Klatt, C. G., Inskeep, W. P., Herrgård, M. J., Jay, Z. J., Rusch, D. B., Tringe, S. G., et al. (2013). Community structure and function of high-temperature chlorophototrophic microbial mats inhabiting diverse geothermal environments. *Front. Microbiol.* 4:106. doi: 10.3389/fmicb.2013.00106
- Klatt, C. G., Wood, J. M., Rusch, D. B., Bateson, M. M., Hamamura, N., Heidelberg, J. F., et al. (2011). Community ecology of hot spring cyanobacterial mats: Predominant populations and their functional potential. *ISME J.* 5, 1262–1278. doi: 10.1038/ismej.2011.73
- Koonin, E. V., and Krupovic, M. (2015). Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Nat. Rev. Genet.* 16, 184–192. doi: 10.1038/nrg3859
- Koonin, E. V., and Makarova, K. S. (2019). Origins and evolution of CRISPR-Cas systems. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 374:20180087. doi: 10.1098/rstb.2018.0087
- Koonin, E. V., Makarova, K. S., and Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* 37, 67–78. doi: 10.1016/j.mib.2017.05.008
- Koonin, E. V., Makarova, K. S., Wolf, Y. I., and Krupovic, M. (2020). Evolutionary entanglement of mobile genetic elements and host defence systems: Guns for hire. *Nat. Rev. Genet.* 21, 119–131. doi: 10.1038/s41576-019-0172-9
- Krishnan, A., Burroughs, A. M., Iyer, L. M., and Aravind, L. (2020). Comprehensive classification of ABC ATPases and their functional radiation in nucleoprotein dynamics and biological conflict systems. *Nucleic Acids Res.* 48, 10045–10075. doi: 10.1093/nar/gkaa726
- Krupovic, M., and Koonin, E. V. (2016). Self-synthesizing transposons: Unexpected key players in the evolution of viruses and defense systems. *Curr. Opin. Microbiol.* 31, 25–33. doi: 10.1016/j.mib.2016.01.006
- Krupovic, M., Béguin, P., and Koonin, E. V. (2017). Casposons: Mobile genetic elements that gave rise to the CRISPR-Cas adaptation machinery. *Curr. Opin. Microbiol.* 38, 36–43. doi: 10.1016/j.mib.2017.04.004
- Krupovic, M., Makarova, K. S., Forterre, P., Prangishvili, D., and Koonin, E. V. (2014). Casposons: A new superfamily of self-synthesizing DNA transposons at the origin of prokaryotic CRISPR-Cas immunity. *BMC Biol.* 12:36. doi: 10.1186/1741-7007-12-36
- Krupovic, M., Shmakov, S., Makarova, K. S., Forterre, P., and Koonin, E. V. (2016). Recent mobility of casposons, self-synthesizing transposons at the origin of the CRISPR-cas immunity. *Genome Biol. Evol.* 8, 375–386. doi: 10.1093/gbe/evw006
- Kunin, V., He, S., Warnecke, F., Peterson, S. B., Garcia Martin, H., Haynes, M., et al. (2008). A bacterial metapopulation adapts locally to phage predation despite global dispersal. *Genome Res.* 18, 293–297. doi: 10.1101/gr.6835308
- Landsberger, M., Gandon, S., Meaden, S., Rollie, C., Chevallereau, A., Chabas, H., et al. (2018). Anti-CRISPR Phages Cooperate to Overcome CRISPR-Cas Immunity. *Cell* 174, 908–916.e12. doi: 10.1016/j.cell.2018.05.058
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Le, S. Q., and Gascuel, O. (2008). An improved general amino acid replacement matrix. *Mol. Biol. Evol.* 25, 1307–1320. doi: 10.1093/molbev/msn067
- Letunic, I., and Bork, P. (2019). Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239
- Li, D., Liu, C. M., Luo, R., Sadakane, K., and Lam, T. W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi: 10.1093/bioinformatics/btv033
- Li, S. J., Hua, Z. S., Huang, L. N., Li, J., Shi, S. H., Chen, L. X., et al. (2014). Microbial communities evolve faster in extreme environments. *Sci. Rep.* 4:6205. doi: 10.1038/srep06205
- López-López, O., Cerdán, M. E., and González-Siso, M. I. (2013). Hot spring metagenomics. *Life* 3, 308–320. doi: 10.3390/life3020308
- Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., et al. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 47, W636–W641. doi: 10.1093/nar/gkz268
- Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., et al. (2011). Evolution and classification of the CRISPR–Cas systems. *Nat. Rev. Microbiol.* 9, 467–477. doi: 10.1038/nrmicro2577
- Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015). An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.* 13, 722–736. doi: 10.1038/nrmicro3569
- Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2013). The basic building blocks and evolution of CRISPR-Cas systems. *Biochem. Soc. Trans.* 41, 1392–1400. doi: 10.1042/BST20130038
- Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., et al. (2020a). Evolutionary classification of CRISPR–Cas systems: A burst of class 2 and derived variants. *Nat. Rev. Microbiol.* 18, 67–83. doi: 10.1038/s41579-019-0299-x
- Makarova, K. S., Wolf, Y. I., Shmakov, S. A., Liu, Y., Li, M., and Koonin, E. V. (2020b). Unprecedented diversity of unique CRISPR-Cas-Related Systems and Cas1 Homologs in Asgard Archaea. *Cris. J.* 3, 156–163. doi: 10.1089/crispr.2020.0012
- Marchler-Bauer, A., and Bryant, S. H. (2004). CD-Search: Protein domain annotations on the fly. *Nucleic Acids Res.* 32, W327–W331. doi: 10.1093/nar/gkh454
- Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., et al. (2017). CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 45, D200–D203. doi: 10.1093/nar/gkw1129
- Marchler-Bauer, A., Panchenko, A. R., Shoemaker, B. A., Thiessen, P. A., Geer, L. Y., and Bryant, S. H. (2002). CDD: A database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res.* 30, 281–283. doi: 10.1093/nar/30.1.281
- Massello, F. L., Chan, C. S., Chan, K. G., Goh, K. M., Donati, E., and Urbiet, M. S. (2020). Meta-analysis of microbial communities in hot springs: Recurrent taxa and complex shaping factors beyond pH and temperature. *Microorganisms* 8:906. doi: 10.3390/microorganisms8060906
- McGinn, J., and Marraffini, L. A. (2019). Molecular mechanisms of CRISPR–Cas spacer acquisition. *Nat. Rev. Microbiol.* 17, 7–12. doi: 10.1038/s41579-018-0071-7
- Meaden, S., Capria, L., Alseth, E., Gandon, S., Biswas, A., Lenzi, L., et al. (2021). Phage gene expression and host responses lead to infection-dependent costs of CRISPR immunity. *ISME J.* 15, 534–544. doi: 10.1038/s41396-020-00794-w
- Merino, N., Aronson, H. S., Bojanova, D. P., Feyhl-Buska, J., Wong, M. L., Zhang, S., et al. (2019). Living at the extremes: Extremophiles and the limits of life in a planetary context. *Front. Microbiol.* 10:780. doi: 10.3389/fmicb.2019.0780
- Meyer-Dombard, D. R., Shock, E. L., and Amend, J. P. (2005). Archaeal and bacterial communities in geochemically diverse hot springs of Yellowstone National Park, USA. *Geobiology* 3, 211–227. doi: 10.1111/j.1472-4669.2005.00052.x
- Mohanraju, P., Makarova, K. S., Zetsche, B., Zhang, F., Koonin, E. V., and Van Der Oost, J. (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* 353, aad5147. doi: 10.1126/science.aad5147
- Mohanraju, P., Saha, C., van Baaren, P., Louwen, R., Staals, R. H. J., and van der Oost, J. (2022). Alternative functions of CRISPR–Cas systems in the evolutionary arms race. *Nat. Rev. Microbiol.* 20, 351–364. doi: 10.1038/s41579-021-00663-z
- Moya-Beltrán, A., Makarova, K. S., Acuña, L. G., Wolf, Y. I., Covarrubias, P. C., Shmakov, S. A., et al. (2021). Evolution of Type IV CRISPR-Cas systems: Insights from CRISPR loci in integrative conjugative elements of Acidithiobacillia. *Cris. J.* 4, 656–672. doi: 10.1089/crispr.2021.0051
- Moya-Beltrán, A., Rojas-Villalobos, C., Díaz, M., Guiliani, N., Quatrini, R., and Castro, M. (2019). Nucleotide second messenger-based signaling in extreme acidophiles of the Acidithiobacillus species complex: Partition between the core

- and variable gene complements. *Front. Microbiol.* 10:381. doi: 10.3389/fmicb.2019.00381
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., et al. (2020). *Vegan community ecology package: Ordination methods, diversity analysis and other functions for community and vegetation ecologists*. R Packag. version 2.5-7. Available online at: <http://apps.worldagroforestry.org/publication/vegan-community-ecology-package-ordination-methods-diversity-analysis-and-other> (accessed May 21, 2021).
- Parmar, K., Dafale, N., Pal, R., Tikariha, H., and Purohit, H. (2018). An insight into phage diversity at environmental habitats using comparative metagenomics approach. *Curr. Microbiol.* 75, 132–141. doi: 10.1007/s00284-017-1357-0
- Pericard, P., Dufresne, Y., Couderc, L., Blanquart, S., and Touzet, H. (2018). MATAM: Reconstruction of phylogenetic marker genes from short sequencing reads in metagenomes. *Bioinformatics* 34, 585–591. doi: 10.1093/bioinformatics/btx644
- Pinilla-Redondo, R., Mayo-Muñoz, D., Russel, J., Garrett, R. A., Randau, L., Sørensen, S. J., et al. (2020). Type IV CRISPR–Cas systems are highly diverse and involved in competition between plasmids. *Nucleic Acids Res.* 48, 2000–2012. doi: 10.1093/NAR/GKZ1197
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219
- Russel, J., Pinilla-Redondo, R., Mayo-Muñoz, D., Shah, S. A., and Sørensen, S. J. (2020). CRISPRCasTyper: Automated identification, annotation, and classification of CRISPR–Cas Loci. *Cris. J.* 3, 462–469. doi: 10.1089/crispr.2020.0059
- Sampson, T. R., and Weiss, D. S. (2013). Alternative Roles for CRISPR/Cas Systems in Bacterial Pathogenesis. *PLoS Pathog.* 9:e1003621. doi: 10.1371/JOURNAL.PPAT.1003621
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Sharp, C. E., Brady, A. L., Sharp, G. H., Grasby, S. E., Stott, M. B., and Dunfield, P. F. (2014). Humboldt's spa: Microbial diversity is controlled by temperature in geothermal environments. *ISME J.* 8, 1166–1174. doi: 10.1038/ismej.2013.237
- Shmakov, S. A., Sitnik, V., Makarova, K. S., Wolf, Y. I., Severinov, K. V., and Koonin, E. V. (2017a). The CRISPR spacer space is dominated by sequences from species-specific mobilomes. *MBio* 8:e01397-17. doi: 10.1128/mBio.01397-17
- Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., et al. (2017b). Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.* 15, 169–182. doi: 10.1038/nrmicro.2016.184
- Silas, S., Makarova, K. S., Shmakov, S., Páez-Espino, D., Mohr, G., Liu, Y., et al. (2017). On the origin of reverse transcriptase- using CRISPR–Cas systems and their hyperdiverse, enigmatic spacer repertoires. *MBio* 8:e00897-17.
- Steinegger, M., and Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* 35, 1026–1028. doi: 10.1038/nbt.3988
- Strazzulli, A., Fusco, S., Cobucci-Ponzano, B., Moracci, M., and Contursi, P. (2017). Metagenomics of microbial and viral life in terrestrial geothermal environments. *Rev. Environ. Sci. Biotechnol.* 16, 425–454. doi: 10.1007/s11157-017-9435-0
- Sullivan, M. J., Petty, N. K., and Beatson, S. A. (2011). Easyfig: A genome comparison visualizer. *Bioinformatics* 27, 1009–1010. doi: 10.1093/BIOINFORMATICS/BTR039
- Taylor, H. N., Warner, E. E., Armbrust, M. J., Crowley, V. M., Olsen, K. J., and Jackson, R. N. (2019). Structural basis of Type IV CRISPR RNA biogenesis by a Cas6 endoribonuclease. *RNA Biol.* 16, 1438–1447. doi: 10.1080/15476286.2019.1634965
- Tian, Y., Liu, R. R., Xian, W. D., Xiong, M., Xiao, M., and Li, W. J. (2020). A novel thermal Cas12b from a hot spring bacterium with high target mismatch tolerance and robust DNA cleavage efficiency. *Int. J. Biol. Macromol.* 147, 376–384. doi: 10.1016/j.jbiomac.2020.01.079
- Tyson, G. W., and Banfield, J. F. (2008). Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ. Microbiol.* 10, 200–207. doi: 10.1111/j.1462-2920.2007.01444.x
- Vale, P. F., Lafforgue, G., Gatchitch, F., Gardan, R., Moineau, S., and Gandon, S. (2015). Costs of CRISPR–Cas-mediated resistance in *Streptococcus thermophilus*. *Proc. R. Soc. B* 282:20151270. doi: 10.1098/rspb.2015.1270
- Van Houte, S., Ekroth, A. K. E., Broniewski, J. M., Chabas, H., Ashby, B., Bondy-Denomy, J., et al. (2016). The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* 532, 385–388. doi: 10.1038/nature17436
- Weinberger, A. D., Wolf, Y. I., Lobkovsky, A. E., Gilmore, M. S., and Koonin, E. V. (2012). Viral diversity threshold for adaptive immunity in prokaryotes. *MBio* 3:e00456-12. doi: 10.1128/mBio.00456-12
- Weissman, J. L., Laljani, R. M. R., Fagan, W. F., and Johnson, P. L. F. (2019). Visualization and prediction of CRISPR incidence in microbial trait-space to identify drivers of antiviral immune strategy. *ISME J.* 13, 2589–2602. doi: 10.1038/s41396-019-0411-2
- Westra, E. R., Dowling, A. J., Broniewski, J. M., and van Houte, S. (2016). Evolution and Ecology of CRISPR. *Annu. Rev. Ecol. Evol. Syst.* 47, 307–331. doi: 10.1146/annurev-ecolsys-121415-032428
- Westra, E. R., Van Houte, S., Oyesiku-Blakemore, S., Makin, B., Broniewski, J. M., Best, A., et al. (2015). Parasite exposure drives selective evolution of constitutive versus inducible defense. *Curr. Biol.* 25, 1043–1049. doi: 10.1016/j.cub.2015.01.065
- Wu, R., Chai, B., Cole, J. R., Gunturu, S. K., Guo, X., Tian, R., et al. (2020). Targeted assemblies of cas1 suggest CRISPR–Cas's response to soil warming. *ISME J.* 14, 1651–1662. doi: 10.1038/s41396-020-0635-1
- Zablocki, O., van Zyl, L., and Trindade, M. (2018). Biogeography and taxonomic overview of terrestrial hot spring thermophilic phages. *Extremophiles* 22, 827–837. doi: 10.1007/s00792-018-1052-5